

Transition path sampling: throwing ropes over mountains in the dark

Peter G Bolhuis[†], Christoph Dellago[‡], Phillip L Geissler[‡] and David Chandler[‡]

[†] Department of Chemistry, Lensfield Rd, Cambridge CB2 1EW, UK

[‡] Department of Chemistry, University of California, Berkeley, CA 94720, USA

Received 5 October 1999

Abstract. Understanding rare transitions occurring in complex systems, for instance chemical reactions in solution, poses the problem of finding and analysing the trajectories that move from one basin of attraction to another on a complicated potential energy surface. We have developed a systematic approach for finding these trajectories using computer simulations without preconceived knowledge of transition states. The approach follows from a novel statistical mechanics and thermodynamics of trajectories and has been demonstrated with several applications.

(Some figures in this article appear in black and white in the printed version.)

1. Introduction

We have developed a general computational method, called transition path sampling, for finding rare pathways connecting stable states in complex systems [1–5]. This method allows one to study, for instance, chemical reactions in solution or conformational transitions of large biopolymers. The algorithm requires no preconceived notion of mechanisms or transition states. Transition path sampling is metaphorically akin to throwing ropes over rough mountain passes in the dark. ‘Throwing ropes’ in the sense that one shoots short trajectories, attempting to reach one stable state from another. ‘In the dark’ because the high-dimensional systems are so complex that it is generally literally impossible to visualize the topography of relevant energy surfaces (‘mountains’). In such cases, it is unlikely that the first throw of the rope will be successful, but one can learn from failures, and there should be an optimum algorithm, i.e., sequence of throws, with which success is obtained efficiently. We have developed this type of sequence, and we believe that it opens the way for many heretofore impossible computational studies of the dynamic pathways of chemical reactions in clusters and in condensed phases.

2. Rare but important events in complex systems

Often, dynamical processes of interest occur on timescales that are very long compared to the molecular timescale. For example, the dissociation of a weak acid in water might occur with a half-life of a millisecond, while elementary steps of molecular motions in water occur on a timescale of femtoseconds. Similarly, timescales for folding for the smallest of proteins are in the range of milliseconds to microseconds, while those for small-amplitude motions

of amino acid side chains and water solvent are again femtoseconds. This wide disparity of timescales can present serious computational challenges when employing straightforward molecular simulation.

One way around this problem is to focus on the dynamical bottleneck for the rare event—the transition state. If its location is known, one may first move the system reversibly to the transition state and then initiate many fleeting trajectories from that state [6]. The first step determines the reversible work and thus the probability for observing the system at the transition state. The subsequent trajectories determine the probability for successfully crossing the threshold. Together, they give the rate for the rare event. While theoretically sound, this two-step procedure is limited in applicability because it presupposes knowledge of the transition state. In most interesting cases, transition states are not or poorly known.

For low-dimensional systems involving a small number of atoms, transition states can be located numerically with various algorithms that systematically search for saddle points in the potential energy surface. For higher-dimensional systems, however, the potential energy surface will typically contain many saddle points. Explicit enumeration of saddle points is feasible for a cluster of ten or fewer atoms, but provides no means to distinguish saddle points that are dynamically irrelevant from those that are dynamically relevant. For a chaotic system—large polyatomic molecules, large clusters and condensed phases—potential energy surfaces are rough on the scale of thermal energies, $k_B T$, and dense in saddle points. Effectively, therefore, there is generally a practically uncountable number of transition states. Searching for a few such states is insufficient. Instead, one wants to locate and sample an ensemble of transition states. Transition path sampling accomplishes this task.

3. Transition path sampling

Transition path sampling is based on a generalization of standard Monte Carlo simulation procedures. In its standard form, a Monte Carlo calculation performs a random walk in configuration space. The walk is biased to ensure that the most important regions of configuration space are adequately sampled. Specifically, in a Monte Carlo random walk, a configuration x is visited in proportion to its probability $p(x)$. The walk may be initiated far from a typical configuration, but after some equilibration period, the bias drives the system to the important regions of configuration space. This feature, crucial to its success, is called ‘importance sampling’.

Importance sampling can be generalized to trajectory space. Consider the ensemble of all dynamical trajectories of a fixed duration, say 1 ps. Most of these trajectories will be localized near some basin of attraction, i.e. a long-lived collection of neighbouring microstates called a *stable state*. Rare crossings between stable states will comprise a small subset of these 1 ps trajectories. For example, if the process of interest occurs roughly once every millisecond, then only one out of a billion trajectories 1 ps long will exemplify that process. Transition path sampling harvests members of such rare subensembles.

Suppose the rare processes of interest are transitions between states A and B, where A and B are characterized by their respective population operators, $h_A(x)$ and $h_B(x)$. Here, x denotes a point in phase space. The probability weight for such unusual trajectories is

$$P(x_0) \propto \rho(x_0) h_A(x_0) h_B(x_t) \quad (1)$$

where $\rho(x_0)$ is the unconstrained distribution of initial phase-space points, x_0 , and x_t is the phase-space point at a time t later. (For simplicity, but not of necessity, we might be considering deterministic dynamics, in which case x_t follows from the initial phase-space point x_0 by integration of the equations of motion.) The population operators $h_A(x)$ and $h_B(x)$

are used to select the reactive trajectories from the set of all possible trajectories. Therefore, the distribution $P(x)$ is a statistical description of all trajectories connecting A with B. We call this set of reactive paths the transition path ensemble. Transition path sampling is performed by carrying out a random walk in trajectory space, biased according to the path distribution $P(x)$ [2–5]. In particular, the shooting algorithm depicted schematically in figure 1 provides an efficient method for harvesting trajectories with weight $P(x)$.

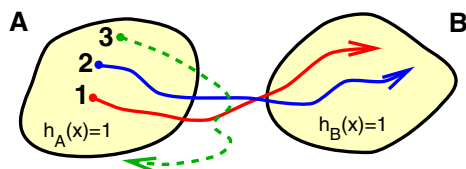


Figure 1. In a transition path sampling simulation the path ensemble is sampled by generating a random walk in path space. This figure shows two steps of a such a random walk. Here, path 2 is generated by slightly modifying path 1. Since path 2 is still connecting A and B it is accepted as the new path. By the same procedure path 3 is generated from path 2. Now, however, the new path is not a reactive trajectory. Therefore it is rejected. Indeed, it is this sequence of acceptances and rejections which ensures that the correct path ensemble is sampled in our computer simulation.

In this algorithm, one needs only to characterize stable or long-lived states. Nothing is declared about the dynamical pathways (i.e., trajectories) that join these states.

The time-correlation function that determines the rate constant for $A \rightarrow B$ transitions can be viewed as a ratio of partition functions for two different transition path ensembles. In particular,

$$\frac{\langle h_A(0)h_B(t) \rangle}{\langle h_A \rangle} = \left(\sum_{[x(t)]} \rho(x_0) h_A(x_0) h_B(x_t) \right) / \left(\sum_{[x(t)]} \rho(x_0) h_A(x_0) \right) \quad (2)$$

where $\sum_{[x(t)]}$ denotes the sum over all trajectories $x(t)$ generated by the dynamics of interest[†]. In view of equation (2), one sees that the calculation of time-correlation functions is isomorphic with the calculation of reversible work or free-energy differences [5]. Thus, from this perspective, standard tools of equilibrium computer simulation, like thermodynamic perturbation theory, can be adopted to compute dynamical properties, like rate constants. In the next section we present four examples of the application of the transition path sampling method.

4. Applications

4.1. Rearrangement processes in Lennard-Jones clusters

Even simple clusters of a few atoms can show interesting dynamical behaviour. We studied rearrangement processes in a cluster consisting of seven Lennard-Jones particles in two dimensions [3]. Using a quenching technique in path space we were able to identify the transition mechanisms without prior knowledge of transition states. A typical high-probability path is shown in figure 2. The isomorphism between time-correlation functions and free energies was employed to calculate rate constants for this reorganization process.

[†] For deterministic trajectories, this sum is accomplished by summing over initial conditions x_0 . For non-deterministic stochastic dynamics the sum must weight different realizations of the random force consistent with the probability distribution for those forces.

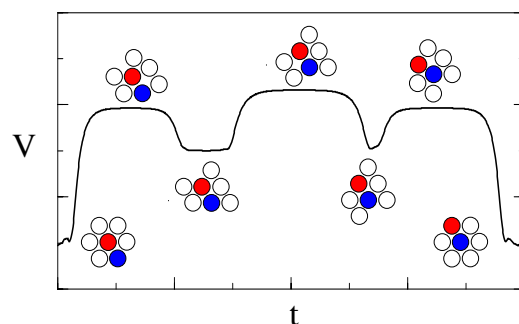


Figure 2. The potential energy profile as a function of time obtained by transition path sampling for the isomerization of a seven-particle Lennard-Jones cluster. In the course of the transition the grey particle initially at the centre migrates to its surface and is replaced by the black particle at the centre of the cluster. The configurations depicted below the curves are the stable states visited during the transition and the configurations above the curves are the transition states.

4.2. Chemical dynamics of the protonated water trimer

We have investigated proton-transfer dynamics of a protonated water trimer using transition path sampling. Transfer of a proton from one water molecule to another involves significant rearrangement of the cluster. Figure 3(a) shows snapshots along a typical transition path. During the transition, water molecule 3, which initially accepts a hydrogen bond from molecule 1, travels to the opposite side of the cluster while the system crosses a region close to a saddle point on the potential energy surface. This rearrangement of the water molecules is accompanied by a proton transfer from molecule 1 to molecule 2. Two different transition state regions are important for the transition dynamics. They are shown in figure 3(b).

Since the transferring proton is a light particle, one might suspect that quantum mechanical effects are important for the transfer of the proton from one water to another. More specifically, one could imagine that to move from the donating water molecule to the accepting water molecule the proton tunnels through the potential energy barrier. By studying reactive trajectories [7] we discovered that tunnelling plays only a minor role in this proton-transfer process (except at very low temperature). As mentioned above, proton transfer in this system is driven by the rearrangement of the water molecules. As a consequence the transferring proton feels a monostable potential trough the whole transition. Therefore no tunnelling is necessary to transport the proton from one water molecule to the other.

4.3. Ion pair dissociation in water

Consider an ion pair, for instance Na^+ and Cl^- , solvated in water and initially in contact. Though the ions attract each other the associated state is only metastable and will eventually decay due to thermal fluctuations. An average separation time of the order of 15 ps indicates that the system has to surmount a barrier of several $k_B T$ in order to initiate the separation of the ion pair. Until recently it was widely believed that the solvating water molecules, while acting as a polarizable continuum favouring the separation and providing the energy to escape over the barrier, do not have a more specific role related to their molecular nature. In contrast, transition path sampling has demonstrated [8] that the dynamics of the solvent is important for the dissociation process. In other words, a reaction coordinate that correctly describes the mechanism of ion dissociation must include components involving the solvent coordinates.

Specifically, the addition of a water molecule to the shell of solvent surrounding the Na^+

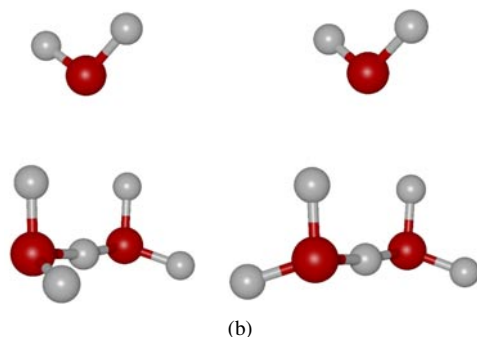
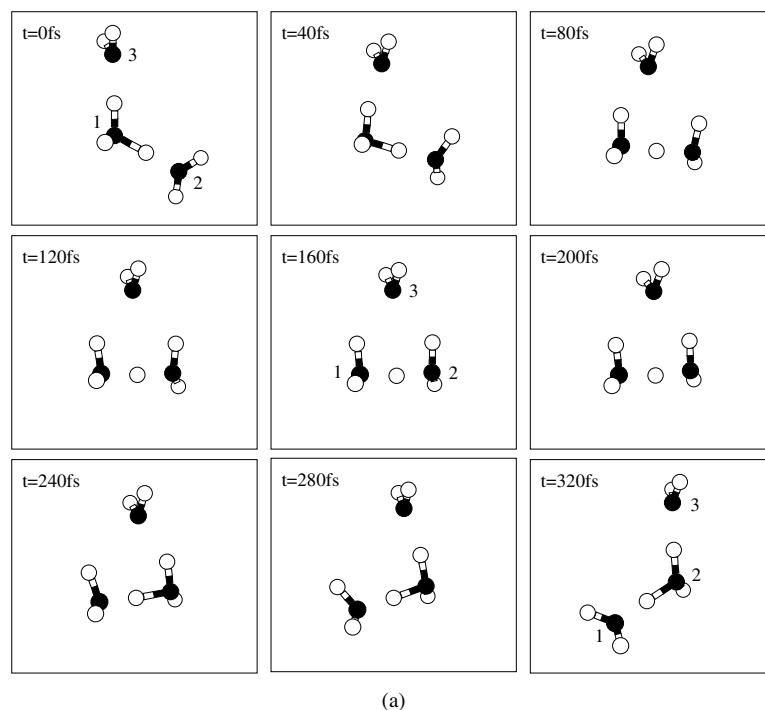


Figure 3. (a) Snapshots of the protonated water trimer along a reactive trajectory. During the transition a proton is transferred from water molecule 1 to water molecule 2 as molecule 3 moves from the left to the right side of the cluster. (b) A view of the two relevant saddle points found using the path sampling procedure. In the higher-energy transition state (left-hand side) the out-of-plane hydrogen atoms of water molecules 1 and 2 lie on the same side of the oxygen plane. In the lower-energy transition state (right-hand side) these hydrogens lie on opposite sides of the plane, so the dipoles of water molecules 1 and 2 are antiparallel.

ion with the associated rearrangement of the solvent molecules strongly assists the dissociation of the ions.

4.4. Conformational change in alanine dipeptide

Transition path sampling can also be applied to conformational changes in biomolecules, such as protein folding. Such processes are notoriously difficult to assess, due to a lack of a proper description of the reaction coordinate. To demonstrate the feasibility of this kind

of application, we studied the transition from the C_{eq} - to the α_R -state of alanine dipeptide in water [9]. The transition state ensemble was found to be bimodal, indicating there are two pathways which are both automatically visited by transition path sampling. Furthermore, the distribution of transition states is extremely broad. Transition states can be quite far away from the maximum of the potential of mean force. The use of a potential of force to find transition states is therefore clearly misleading. Rather, the conformational transition is dominated by diffusion on an almost flat free-energy barrier. Furthermore, the transition state ensemble is very broad in a fashion indicating that water molecules surrounding the dipeptide make the energy landscape rough and that the coordinates of the water molecules play a direct, and not simply implicit, role in the isomerization pathway.

5. Conclusions

The transition path sampling method enables the computational scientist to study processes involving rare events in highly complicated systems where traditional methods fail. This method does not require knowledge of the location of dynamical bottlenecks, but rather can be used to find them. Furthermore, the method can be easily blended with any procedure that generates reversible trajectories. For example, the method has been combined with Car and Parrinello's *ab initio* molecular dynamics and with molecular dynamics packages for biomolecules such as AMBER and CHARMM. Transition path sampling is therefore an ideal tool for the exploration of the dynamics of complex systems in physics, chemistry, and biology.

Acknowledgments

This work was initiated with support from the National Science Foundation and completed with support from the Department of Energy through the Chemical Sciences Division of Lawrence Berkeley National Laboratory.

References

- [1] Chandler D 1998 Finding transition pathways: throwing ropes over rough mountain passes, in the dark *Computer Simulation of Rare Events and Dynamics of Classical and Quantum Condensed-Phase Systems—Classical and Quantum Dynamics in Condensed Phase Simulations* ed B J Berne, G Ciccotti and D F Coker (Singapore: World Scientific) pp 51–66
- [2] Dellago C, Bolhuis P G, Csajka F S and Chandler D 1998 *J. Chem. Phys.* **108** 1964
- [3] Dellago C, Bolhuis P G and Chandler D 1998 *J. Chem. Phys.* **108** 9236
- [4] Bolhuis P G, Dellago C and Chandler D 1998 *Faraday Discuss. Chem. Soc.* **110** 421
- [5] Dellago C, Bolhuis P G and Chandler D 1999 *J. Chem. Phys.* **110** 6617
- [6] This two-step procedure was pioneered by Charles Bennett and David Chandler. See, for example, Bennett C H 1977 *Algorithms for Chemical Computations (ACS Symp. Ser. No 46)* ed R E Christofferson (Washington, DC: American Chemical Society) p 63
Chandler D 1978 *J. Chem. Phys.* **68** 2959
Related and earlier work has been discussed in a brief history communicated by Chandler D 1988 *Faraday Discuss. Chem. Soc.* **85** 341
- [7] Geissler P, Dellago C and Chandler D 1999 *Phys. Chem. Chem. Phys.* **1** 1317
- [8] Geissler P, Dellago C and Chandler D 1999 *J. Phys. Chem. B* **103** 3706
- [9] Bolhuis P G, Dellago C and Chandler D 2000 to be published